DOCUMENT RESUME

ED 425 598 EC 306 939

| | |
|---|---|
| AUTHOR | Seyfarth, Allison; Ysseldyke, James E.; Thurlow, Martha L. |
| TITLE | An Analysis of Perceived Desirability, Feasibility, and Actual Use of Specific Criteria for Large-Scale Assessment and Accountability Systems. Technical Report 21. |
| INSTITUTION | National Center on Educational Outcomes, Minneapolis, MN.; Council of Chief State School Officers, Washington, DC.; National Association of State Directors of Special Education, Alexandria, VA. |
| SPONS AGENCY | Special Education Programs (ED/OSERS), Washington, DC. |
| PUB DATE | 1998-07-00 |
| NOTE | 39p. |
| CONTRACT | H159C50004 |
| AVAILABLE FROM | National Center on Education Outcomes, University of Minnesota, 350 Elliott Hall, 75 East River Road, Minneapolis, MN 55455; telephone: 612-624-8561; fax: 612-624-0879; Web site: http://www.coled.umn.edu/NCEO ($10). |
| PUB TYPE | Reports - Research (143) |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | *Accountability; *Disabilities; *Educational Assessment; Elementary Secondary Education; *Evaluation Criteria; Input Output Analysis; Outcomes of Education; Recordkeeping; School Effectiveness; Special Education; *Student Participation; *Teacher Attitudes; Teacher Participation |
| IDENTIFIERS | Alternative Assessment; *Testing Accommodations (Disabilities) |

ABSTRACT

In the mid 1990s, the National Center on Educational Outcomes worked with stakeholder groups to establish a set of criteria for use in large-scale assessments. The criteria addressed issues of the participation of students with disabilities in the assessments, assessment accommodations, and the reporting of results for students with disabilities. This study examined the extent to which the criteria were perceived to be desirable and feasible. A national sample of 222 administrators, teachers, and other Individualized Education Program team members rated the criteria for desirability and feasibility. The average rating of desirability for the criteria was high, with three main exceptions: (1) specific criteria that were seen as less desirable, such as global statements about including all students in assessments; (2) statements that were seen as less feasible than others included those rated low in desirability, as well as statements that called for parent understanding of the issues and record-keeping of accommodations provided; and (3) all criteria but one were rated as significantly less feasible than desirable. Criteria rated high on actual use included those focused on making individual decisions about students or on documentation of the decisions. (Author/CR)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

Technical Report 21

ED 425 598

# An Analysis of Perceived Desirability, Feasibility, and Actual Use of Specific Criteria for Large-Scale Assessment and Accountability Systems

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Ysseldyke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

NATIONAL CENTER ON EDUCATIONAL OUTCOMES

*In collaboration with:*

Council of Chief State School Officers (CCSSO)

National Association of State Directors of Special Education (NASDSE)

EC 306537

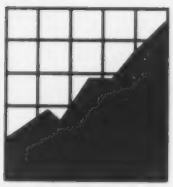
2

Technical Report 21

# An Analysis of Perceived Desirability, Feasibility, and Actual Use of Specific Criteria for Large-Scale Assessment and Accountability Systems

Allison Seyfarth • James E. Ysseldyke • Martha L. Thurlow

July, 1998



NATIONAL CENTER ON EDUCATIONAL OUTCOMES

The Center is supported through a Cooperative Agreement (#H159C50004) with the Division of Innovation and Development, Office of Special Education Programs, U.S. Department of Education. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.

**NCEO Core Staff**

Robert H. Bruininks
Judith L. Elliott
Michael L. Moore
Dorene L. Scott
Sandra J. Thompson
Martha L. Thurlow, Associate Director
James E. Ysseldyke, Director

Additional copies of this document may be ordered for $10.00 from:

National Center on Educational Outcomes
University of Minnesota • 350 Elliott Hall
75 East River Road • Minneapolis, MN 55455
Phone 612/624-8561 • Fax 612/624-0879
http://www.coled.umn.edu/NCEO

## Executive Summary

In the mid 1990s the National Center on Educational Outcomes (NCEO) worked with stakeholder groups to establish a set of criteria for use in large-scale assessments. The criteria addressed issues of the participation of students with disabilities in the assessments, accommodations to the assessments, and the reporting of results for students with disabilities. We examined the extent to which the criteria were perceived to be desirable and feasible. A national sample of 222 administrators, teachers, and other IEP team members rated the criteria for desirability and feasibility. The average rating of desirability for the criteria was high, but there were three clear outliers: (1) specific criteria that were seen as less desirable than the others included global statements about including all students in assessments; (2) statements that were seen as less feasible than the others included those rated low in desirability, as well as those that called for parent understanding of the issues and record-keeping of accommodations provided; and (3) all criteria but one were rated to be significantly less feasible than desirable. Criteria rated high on actual use included those focused on making individual decisions for students or on documentation of the decisions.

From this survey, it appears that educators agree that implementing guidelines to increase the participation of students with disabilities will be somewhat difficult. Teachers and other educators actually implementing the changes will need assistance in doing so in a feasible manner, and will need assistance in helping parents understand the accountability systems and their child's place within those systems. Finally, from this survey, it appears that teachers will more likely accept changes when they have some control over some of the decision-making for their students.

## Table of Contents


Overview ..... 1
- Participation ..... 1
  - Current Status of States' Participation Policies ..... 3
- Accommodations ..... 4
  - Empirical Research on Accommodations ..... 5
  - Current Status of States' Accommodation Policies ..... 6
- Reporting ..... 6
- Criteria ..... 7
- Stakeholders ..... 8
- Implications ..... 8

Methods ..... 9
- Participants ..... 9
- Survey ..... 9
- Procedure ..... 10

Results ..... 11
- Desirability ..... 11
- Feasibility ..... 15
- Comparison of Desirability and Feasibility ..... 15
- Actual Use ..... 16

Discussion ..... 17
- Policy Implications and Future Directions ..... 20

References ..... 22

Appendix A: Markers for the Participation, Accommodation, and Reporting Criteria ..... 25

Appendix B: Copy of Survey Instrument ..... 29

## Overview

Statewide assessments are commonplace across the United States. They are usually tests or other performance measures intended to document the educational achievement of the students in a state. In 1996, 48 states reported that they either had a state assessment in place or were developing one. Iowa and Wyoming were the only states without a statewide assessment in development or in place (Bond, Braskamp, & Roeber, 1996), and Wyoming is now starting to develop a statewide assessment.

As statewide assessments become more common, their complexities become more obvious. Some of this complexity involves *what* skill will be tested even when the content area is the same. For example, in a reading test, assessors must decide whether to look at ability to decode or to comprehend text. Another complexity is *how* to test, which involves psychometric issues, as well as how to make modifications or accommodations to tests—or the testing environment—for students with disabilities and students with limited English proficiency. A third complexity is *who* to test, whether to include students with disabilities, and what such assessments mean for those students and their families.

This paper focuses on stakeholder support for decisions on *who* and *how* to test, as well as what to *report* from those tests. There now exists a body of literature addressing policy and research on participation, accommodations, and reporting. This literature forms the basis for criteria for making decisions about these topics. Such criteria were developed by the National Center on Educational Outcomes (NCEO) to guide decision making (Elliott, Thurlow, & Ysseldyke, 1996), but their value to stakeholders was not evaluated as part of the development process. The purpose of this report is to review the literature that supports the criteria, then to present the findings of a study on stakeholders' views of the desirability, feasibility, and actual use of the criteria.

### Participation

In the past couple years, the interest in state assessments, the participation of students with disabilities in them, and the use of accommodations has increased exponentially. New special interest groups have been set up on this topic, such as the Council of Chief State School Officers' (CCSSO) Special Education State Collaborative on Assessment and Student Standards (SCASS), and the American Educational Research Association's (AERA) Special Interest Group on participation of students with disabilities and limited English proficiency in large-scale assessments. This has allowed a number of states and policy organizations to come together to discuss issues of participation, accommodations, and reporting, and collaborate on some of the complexities involved.

Increased interest in accounting for the performance of students with disabilities is due in part

to new laws going into effect (e.g., Title I, IDEA). The new Individuals with Disabilities Education Act (1997 Amendments) requires the participation of students with disabilities in state and district assessments, with accommodations when needed and appropriate. In addition, federal funding agencies are providing funds to explore ways to increase participation and also to examine the effects of accommodations. Both the Office of Special Education Programs (OSEP) and the Office of Educational Research and Improvement (OERI) have conducted two rounds of funding for projects to address these issues.

Another motivator in this work is states' realization that there are negative consequences for accountability systems that do not include students with disabilities. Among the most commonly motivating negative consequences is the increasing rate of referral to special education that occurs when students with disabilities can be exempted from tests that are seen as high stakes for schools or districts (Allington & McGill-Franzen, 1992; Ysseldyke, Thurlow, McGrew, & Shriner, 1994; Zlatos, 1994).

Allington and McGill-Franzen (1992) examined the incidence of retention, remediation, and special education placement for students in grade three in New York state. The researchers looked at three different types of schools: those with consistently high scores on New York's pupil evaluation program (PEP) tests in grade three, those with consistently low scores on the tests, and those with scores that improved over three testing periods they examined (school years of 1978-79, 1984-85, and 1988-89). Adding a complication, New York changed the information that they reported on schools in 1984-85, because in that year and subsequent testing periods, they decided to exclude the information from students with disabilities in the reports to the public.

When examining the rates of retention, referral to special education, and remediation in the schools across those time periods, Allington and McGill-Franzen (1992) found that the proportion of students placed in special education increased across the time period, from 2.1% annually to 3.5% annually. The number of students who were retained also increased during the target time period. More striking is their finding that the schools that reported increased PEP scores across the time period removed or delayed entry into the data stream more than twice as many students (by placing students in special education or retaining students) as did stable, higher achieving schools. Additionally, when comparing schools with improving PEP scores to schools with consistently high scores, there was a more than three times higher rate of increase of student placement in special education. These findings led Allington and McGill-Franzen to the conclusion that "as New York has increased accountability pressure on schools, schools generally have responded by increasing the number of young children who were retained in grade or identified as handicapped" (p. 410). Based on this and other findings, they argued that accountability systems should allow for no exclusion of students, regardless of their disability status, and that systems should use an age-based method for testing, rather than grade-based.

In a follow-up article on their research, McGill-Franzen and Allington (1993) noted that in making their recommendation for participation of all students in statewide accountability assessments, including students with disabilities, they had two outcomes in mind. First, they believed that the current system that allows for exclusion provides an incentive to label students as disabled, and to give them that label early in their school career. Second, they noted that there are little or no current accountability data for students with disabilities. They saw increasing their participation in statewide testing as a means to begin to rectify this situation, in that schools would then have ongoing data regarding the achievement of students with disabilities.

Zlatos, in a 1994 article, examined the percentage of students tested in various school districts. He found a range from a high in Memphis where 93% of students were tested to a low in Boston where 66% of currently enrolled students were tested. Zlatos also reported on an occurrence in the Oklahoma City Public schools in which there was a large increase in the number of exemptions during the late 1980s and early 1990s. During that time, the test scores increased dramatically, from 39% of students scoring above the national norm in 1987-88 to 53% of students scoring above the national norm in 1991-92. The number of schools on the state's probationary list dropped during the same time period from 21 schools on the list to 1. However, a survey of the 20 schools that had worked their way off the state's probationary list found that two out of every three students in the schools had been exempted from the state tests because of a disability or a problem speaking English.

Researchers at NCEO and other institutions have found that the actual rate of participation of students with disabilities in large-scale assessments varies widely across states, from a high of 100 percent to a low of 0 percent (McDonnell, McLaughlin, & Morison, 1997; Shriner & Thurlow, 1992). This is a concern for a number of reasons. Comparisons across states or districts that have different levels of participation are invalid because they are testing different populations of students. Second, by exempting any students receiving special education services, states are in effect informing schools that they are only held responsible for the learning of general education students.

***Current Status of States' Participation Policies***

Thurlow, Seyfarth, Scott, and Ysseldyke (1997) found that currently 40 of the 50 states have active policies on the participation of students with disabilities in statewide testing. Those states that do not have current policies either have no statewide assessment, are developing or revising their assessment, or are developing or revising their participation policies. Most frequently, IEP teams are used by states to determine whether students with disabilities will participate in assessments, and some policies specifically refer to the role of parents or guardians in making the decision. Examining course content as it compares to the construct measured in the assessment is another variable that many states consider in making the participation decision for students with disabilities.

Thurlow et al. (1997) examined other options that are becoming more widely available in participation decisions, including partial participation (only taking a portion of an assessment) and providing alternate assessments. Partial participation was found to be the most popular alternative with about 40% of states (16 states) that had policies providing this option. Alternate assessments, generally designed for students who are not able to meaningfully participate in regular assessments even with accommodations, were offered by 20% of states with policies (8 states).

## Accommodations

Another critical issue in the participation of students with disabilities in large-scale assessments is making testing accommodations. Accommodations are changes to the test or testing environment intended to remove the effect of the disability from the testing. This could involve giving a student with a visual impairment a Braille version of the test, permitting a student in a wheelchair to use a different kind of table to take the test on, or reading a test to a student who has a learning disability in reading (Thurlow, Ysseldyke, & Silverstein, 1993). States vary widely in the accommodations they offer to students with disabilities in large-scale assessments (Thurlow et al., 1997).

There are a number of accommodations typically allowed by states and other testing administrators (e.g., Educational Testing Service); these can be thought of as falling into four categories (Thurlow et al., 1993). These include changes in the *presentation format*, such as offering a Braille edition, orally reading directions, and using large print editions. Changes in the *response format* could include allowing a child to point to a response, the use of sign language to communicate a response, or using a computer for responding. The *setting* of the tests sometimes is changed for students with disabilities; for example, the test could be administered in their home, or alone in a test carrel. Finally, the *timing* of the test is often altered, providing students with extended time, more breaks, or prolonging the test to last for several days.

There are two primary concerns about the use of accommodations: test validity and reliability (Phillips, 1994). When we change some aspect of the testing situation, it becomes unclear whether we are examining the same construct that the test was designed to measure. To use an extreme example, if a test is measuring reading comprehension, and Sally Student, who has a learning disability, receives an accommodation of having the test read to her, it is unclear whether the test is still measuring some kind of reading comprehension, or whether it is measuring listening ability, short term memory, or a combination of all three. Questions about test reliability are mentioned less frequently in the literature (Phillips, 1993); however, if we are changing the nature of the task, we need to examine the reliability of the test to ensure that test results are consistent across administrations.

Beyond concerns of validity and reliability, the basic issue of when to allow accommodations to tests has become increasingly complex. Phillips (1994) notes that accommodating for students with physical disabilities (e.g., blindness, limited mobility, or a temporary condition like a broken arm) was commonplace for many years. Changes for physical disabilities typically included the removal of physical barriers, but was thought not to affect the content of the test. More recently, many requests for testing accommodations have been due to underlying mental disabilities, such as attention deficit disorder and learning disabilities (Phillips, 1994). Since these disabilities are generally not visible, there is a different need for documentation, as well as questions raised about how to modify tests for these students while continuing to use psychometrically appropriate practices.

In the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1985), the authors concluded that there are significant difficulties in using accommodations when testing individuals with disabilities. They cited the need for more empirical research to investigate the effects of testing modifications. Many other authors also cite the lack of empirical evidence about the effects of accommodations (e.g., Siskind, 1993; Thurlow et al., 1993; Thurlow, Hurley, Spicuzza, & El Sawaf, 1996). Siskind (1993) noted that while there has been some research done for large-scale standardized assessments, such as the Scholastic Aptitude Test (SAT), there has been very little research on testing accommodations for criterion-referenced tests, which are used by many states.

***Empirical Research on Accommodations***

Many studies of accommodations have looked at large standardized tests, such as the SAT or the Graduate Record Examination (GRE). One such study examined item performance for students with disabilities who took the SAT with accommodations as compared to the group of students who took the SAT without any accommodations (Bennett, Rock, & Kaplan, 1987). The disability status was not reported for the students who did not use accommodations in testing, though they are assumed to be non-disabled by the authors. When doing item analysis, the researchers condensed the items into clusters in order to reduce the statistical error. They found that two clusters on the mathematical scale were differentially difficult for visually impaired students taking the Braille version of the SAT, one cluster that included questions using graphics in a multiple choice format, and another that included miscellaneous multiple choice items. Additionally, it was found that the algebra comparisons cluster was unexpectedly *easy* for students with learning disabilities taking the test via cassette, and for hearing-impaired students who took the regular exam with extended time. Finally, when comparing items requiring differential amounts of reading, the researchers found that the hearing impaired students who took the regular exam with extended time found the nonreading cluster to be unexpectedly easy for them.

A survey of students who used accommodations in taking the SAT and the GRE examined a number of issues related to testing and accommodation (Ragosta & Kaplan, 1986). This survey asked the participants whether they found the accommodations useful for them, and whether they used similar accommodations in their college-level testing. Overall, the respondents found the accommodations helpful for them (94% overall). However, they were not often using such accommodations in their college testing. For example, 64% of students with visual impairments took the SAT or GRE using a large-type test. In contrast, only 14% of those students used a large-type test in their college tests. One other finding of note from this survey is that 68% of students who eventually took the SAT with accommodations were initially unaware that such accommodations were available, and 19% of GRE test-takers with accommodations were initially unaware of the availability of accommodations.

Phillips (1994) provides some suggestions, based mainly on legal mandates and precedents, for people making accommodations decisions. One basic suggestion is for all school districts and states to have a policy on accommodations in writing. Additionally, she suggests having one individual (or a small group) responsible for making all accommodation decisions. Phillips notes that it is important to make decisions on accommodations on an individual basis, though individuals with similar circumstances should be treated in a similar manner. Finally, she suggests that people requesting accommodations could be asked to sign a release, in order to be able to "flag" their accommodations on relevant testing records. She suggests this step as a way of creating informed consent of release of information while also allowing psychometricians to report their results in an ethical and accurate manner.

***Current Status of States' Accommodation Policies***

Thurlow et al. (1997) found that 39 of the 50 states have active policies on accommodations in state assessments. Again, those that do not have active policies either have no state assessments, are developing or revising their state assessments, or are developing or revising their accommodations policies. One state (South Dakota) does not have a policy because there are no accommodations allowed in the state assessment.

IEP teams are very important in making these decisions (Thurlow et al., 1997). Just over 50% of the states with policies specifically designate the IEP team to make the decision about what accommodations are needed. Approximately 36% of the states with policies refer to the need to link assessment accommodations to those used in instruction.

## Reporting

Another issue raised by the participation of students with disabilities in large-scale assessment is the reporting of results. Parents and schools sometimes are given access to the results, and sometimes not. In addition, over half of states with policies on reporting of results exclude data

from students with disabilities from accountability reports, and many do not document that exclusion (Thurlow, Scott, & Ysseldyke, 1995; Thurlow, Langenfeld, Nelson, Shin, & Coleman, 1998).

In a survey of state special education directors in 1995, one-third of states reported that they could not identify and disaggregate the scores of students with disabilities within their assessment program data bases (Erickson, Thurlow, Thor, & Seyfarth, 1996). Further, among the states that were able to identify and access achievement results for students with disabilities, roughly half chose not to do so. Two years later (Erickson & Thurlow, 1997), 31 states disaggregated scores, yet only about one-half of those reported the information publicly.

In both the 1995 and 1997 surveys, states indicated that they do not analyze and report achievement data for students with disabilities due to lack of resources and time (Erickson et al., 1996). No state indicated that it lacked technical ability to do the analysis or reporting.

Another confusion in reporting the results from students with disabilities in state assessments is the issue of participation rates. States vary a great deal in how they determine how many of their students participated in assessments. Some states exclude certain students, often those with disabilities, and then determine participation rates by comparing the number of eligible students to the number who took the test (Erickson, Thurlow, & Ysseldyke, 1996). Other states include all students in their calculation of rates of participation (Erickson et al., 1996). A detailed analysis of the complicated issues involved in determining participation rates is provided by Almond, Stieber, and Tindal (1997).

## Criteria

Researchers at NCEO developed a set of guidelines for how to address the issues of participation, accommodations, and reporting for students with disabilities in large-scale assessments (Elliott et al., 1996), and then distributed them for comment to policymakers (state directors of special education and assessment) across the nation. These guidelines are based on three basic criteria:

**Criterion 1**. State or district has appropriate written guidelines for the *participation* of students with disabilities in large-scale assessments used for accountability purposes.

**Criterion 2**. State or district has appropriate written guidelines for the use of *accommodations* by students with disabilities in large-scale assessments for accountability purposes.

**Criterion 3**. State or district has appropriate written guidelines for *reporting* the results of students with disabilities on large-scale assessments used for accountability purposes.

Within each of the criteria, markers were identified against which written documents could be evaluated. The guidelines and markers are intended to direct the development and revision of policies for inclusive assessment and accountability systems. The markers for each criterion are included in Appendix A.

## Stakeholders

Many writers in education indicate the need to involve key stakeholders in decisions to improve current school practices (e.g., Henry, Dickey, & Areson, 1991; Renihan & Renihan, 1994). Henry et al. examined the inclusion of stakeholders in the development of Virginia's Educational Performance Recognition (EPR) Program, a statewide educational-performance monitoring system. They found that the inclusion of stakeholders in the development of the program increased the responsiveness of the developers to the issues, concerns, and interests of multiple stakeholder groups, and stimulated methods and approaches that would not have been used without their input.

Gilman (1994) reports on the use of stakeholder groups to determine a model of outcomes and indicators developed by NCEO for early childhood education. She reports that in the development process, a variety of stakeholder groups were included, in order to ensure that many perspectives were incorporated, especially those from a special education or disability perspective. Gilman found that in her sample those with a special education background expressed more support for special education outcomes or indicators. However, across all groups, there was a great deal of agreement on the desired educational outcomes from early childhood education.

Teachers are critically important educational stakeholders. They are very often the people responsible for implementing educational reform, and communicating with parents and students about the reform. In a study of teachers' perceptions of making accommodations or adaptations to classroom-based tests for students with disabilities, Jayanthi, Epstein, Polloway, and Bursuck (1996) found that teachers were mixed in their perceptions of appropriateness. The teachers reported that they were most willing to provide accommodations that interfered the least with their current practices.

## Implications

Research in the areas of participation, accommodations, and reporting is relatively scarce. While there is an absence of a solid empirical base, states are continuing to make policies that will have wide-ranging and long-lasting implications for students with disabilities. Researchers at NCEO responded to this state of affairs by developing a set of criteria intended to guide state practice until there is more research available. This development has relied, until now, on expert

opinion and policymaker input, with little feedback from those actually implementing the practices: teachers and other IEP members. The need to survey the stakeholders who actually will be implementing the practices is extremely important if the practices are to be widely used and accepted at this point in time.

The purpose of this study is to examine stakeholder agreement with NCEO's criteria for participation, accommodations, and reporting of test results for students with disabilities in large-scale assessment. The specific research questions are:

1. Are there criteria that are seen by stakeholders as significantly less desirable than other criteria?

2. Are there criteria that are seen by stakeholders as significantly less feasible than other criteria?

3. Are there criteria for which stakeholders see a significant difference between the desirability of the criterion and the feasibility of the criterion?

4. Are there criteria that stakeholders actually use in making decisions about participation, accommodations, and reporting?

## Methods

### Participants

Participants in this investigation were educators from 18 states, with two states included from each of the nine census regions. Respondents were from the following states (in order of percent of sample): Oregon, Kansas, California, Florida, Michigan, Iowa, Utah, Colorado, Illinois, Texas, New Jersey, Tennessee, Vermont, Maryland, Arkansas, Connecticut, Pennsylvania, and Kentucky. The roles represented in the sample (N = 222) included administrators (N = 154), teachers (N = 34), and other personnel within the schools (N = 34) (see Table 1). The educators in the sample worked in urban (N = 40), suburban (N = 62), and rural districts (N = 80), and some in a combination of urban, suburban, and rural districts (N = 40) (see Table 1).

### Survey

The survey mailed to all participants was a desirability/feasibility/actual use survey on the criteria developed by NCEO researchers. It consisted of 29 statements that recipients were to rate in three ways, using a 5-point Likert scale (1 = low, 5 = high). The first rating focused on

**Table 1. Demographic Information for Survey Respondents, by Position and District Type**

| | Number of Surveys | Percent of Sample |
|---|---|---|
| **Position** | | |
| Administrator, Special Ed. | 121 | 54.5% |
| Administrator, Assessment | 4 | 1.8% |
| Administrator, General Ed. | 2 | 0.9% |
| Administrator, Unspecified | 27 | 12.2% |
| School Psychologist, Evaluator, or Social Worker | 17 | 7.7% |
| Teacher, Special Ed. | 26 | 11.7% |
| Teacher, General Ed. | 2 | 0.9% |
| Teacher, Unspecified | 6 | 2.7% |
| Consultant | 6 | 2.7% |
| Other | 5 | 2.3% |
| Missing | 6 | 2.7% |
| **District Type** | | |
| Urban | 40 | 18.0% |
| Suburban | 62 | 27.9% |
| Rural | 80 | 36.0% |
| Urban and Suburban | 1 | 0.5% |
| Urban and Rural | 5 | 2.3% |
| Suburban and Rural | 12 | 5.4% |
| Urban, Suburban, and Rural | 12 | 5.4% |
| Other | 4 | 1.8% |
| Missing | 6 | 2.7% |

the desirability of the statement, which was defined as the extent to which the criterion was a good idea. The second rating of the statement was on its perceived feasibility, which was defined as how practical it would be to actually implement the practice in the individual's state or district assessments. The final rating of the item focused on the actual use of the practice, which was defined as the extent to which the respondent's state or district presently was implementing the criterion. Space was made available for comments on the survey at the end of the 29 statements. A copy of the survey instrument is included in Appendix B.

## Procedure

Selection of participants for this investigation began by identifying 18 states, two states from each of the nine census regions. Within the nine census regions, states were selected based on whether they had an active state or district assessment system. If there were multiple states to select from within the census region, selection was based on past history of the state's participation in NCEO activities and whether there was a current state director of special education in that state (at least one state was in transition for that position). A list of at least 10 district-level special education directors was requested from the state director of special education in each

state. If a list containing more than 10 directors was returned, study participants were selected randomly.

A packet of four surveys, four cover letters, and four postage-paid return envelopes was sent in May of 1996 to each of the 10 district-level special education directors. The director was asked to complete one survey, and to pass the other three surveys on to other educators in the district who were involved in Individualized Educational Program (IEP) teams. In July of 1996 another survey was sent to the directors who had not returned their original surveys, along with another letter asking them to complete the survey, a postage-paid return envelope, and a piece of gum. For those states that did not have at least a 50% return rate (N = 5 states) by August, phone calls were made to prompt the directors to return their surveys.

Of the 720 surveys mailed, we received a total of 222 surveys back, a return rate of 30.8%. Of those returned, 92 were from the original district-level administrator recipients and 127 were from educators who received surveys from the administrators (3 unknown). The fewest surveys received from any state was 5, from Kentucky, and the most received from any state was 21, from Oregon.

## Results

Each criterion is listed in Table 2. The table also contains means, standard deviations, and number of respondents for each survey item.

### Desirability

To assess the perceived desirability of the criteria in comparison to one another, one-sample t-tests were conducted and compared to the overall sample mean for the desirability of the criteria (M = 4.49). Because 29 comparisons were conducted, significance level was set at 0.005. This analysis resulted in three criteria that were significantly below the mean for desirability. They were:

- Criterion 1, "All students, including students with disabilities, are to participate in state and district assessments," ($t(218) = -7.75$, $p = 0.000$).
- Criterion 6, "A student must participate in an assessment if the student receives *any* instruction on the content assessed, regardless of where instruction occurs," ($t(210) = -12.54$, $p = 0.000$).

**Table 2. Means, Standard Deviations, and Number of Respondents for each Item***

| | Criteria | Desirability | Feasibility | Actual Use |
|---|---|---|---|---|
| 1. | All students, including students with disabilities, are to participate in state and district assessments. | Mean = 3.78**<br>SD = 1.35<br>N = 219 | Mean = 3.29**<br>SD = 1.24<br>N = 219 | Mean = 2.90<br>SD = 1.27<br>N = 220 |
| 2. | The decision about participation is made by a person (or a group of people) who knows the student. | Mean = 4.84<br>SD = 0.51<br>N = 219 | Mean = 4.52<br>SD = 0.69<br>N = 219 | Mean = 3.91**<br>SD = 1.20<br>N = 220 |
| 3. | The decision about participation is based on the student's current level of functioning and learning characteristics. | Mean = 4.71<br>SD = 0.73<br>N = 215 | Mean = 4.35<br>SD = 0.84<br>N = 216 | Mean = 3.64**<br>SD = 1.26<br>N = 216 |
| 4. | A form is used that lists the variables to consider in making participation decisions. | Mean = 4.41<br>SD = 0.97<br>N = 218 | Mean = 4.14<br>SD = 1.06<br>N = 216 | Mean = 2.26<br>SD = 1.47<br>N = 215 |
| 5. | Reason(s) for exclusion is documented. | Mean = 4.76<br>SD = 0.61<br>N = 217 | Mean = 4.45<br>SD = 0.90<br>N = 216 | Mean = 3.01<br>SD = 1.63<br>N = 215 |
| 6. | A student must participate in an assessment if the student receives *any* instruction on the content assessed, regardless of where instruction occurs. | Mean = 3.21**<br>SD = 1.48<br>N = 211 | Mean = 3.03**<br>SD = 1.34<br>N = 211 | Mean = 2.40<br>SD = 1.31<br>N = 208 |
| 7. | Decision about participation is *not* based on the program setting. | Mean = 4.39<br>SD = 1.08<br>N = 218 | Mean = 4.00<br>SD = 1.14<br>N = 216** | Mean = 3.21<br>SD = 1.38<br>N = 214 |
| 8. | Decision about participation is *not* based on the category of disability. | Mean = 4.38<br>SD = 1.14<br>N = 219 | Mean = 4.10<br>SD = 1.14<br>N = 218** | Mean = 3.26<br>SD = 1.37<br>N = 217 |
| 9. | Decision about participation is *not* based on the percentage of time in the mainstream classroom. | Mean = 4.32<br>SD = 1.16<br>N = 219 | Mean = 4.08<br>N = 1.16<br>N = 217 | Mean = 3.22**<br>N = 1.44<br>N = 215 |
| 10. | Decision guidelines allow for some students to participate in an alternate assessment or, when appropriate, in part of an assessment or assessment procedure. | Mean = 4.68<br>SD = 0.72<br>N = 218 | Mean = 4.02<br>SD = 1.12<br>N = 218 | Mean = 2.89<br>SD = 1.46<br>N = 218 |
| 11. | Decision guidelines recognize that only a small percentage of students with disabilities need to participate in an alternate assessment (e.g., those with severe disabilities, <1% of all students). | Mean = 4.11**<br>SD = 1.15<br>N = 213 | Mean = 3.85<br>SD = 1.15<br>N = 210 | Mean = 2.68<br>SD = 1.39<br>N = 212 |

**Table 2. Means, Standard Deviations, and Number of Respondents for each Item (continued)**

| | Criteria | Desirability | Feasibility | Actual Use |
|---|---|---|---|---|
| 12. | Parents understand the state/district accountability system. | Mean = 4.70<br>SD = 0.71<br>N = 218 | Mean = 3.47**<br>SD = 1.06<br>N = 218 | Mean = 2.08<br>SD = 1.02<br>N = 218 |
| 13. | Parents understand participation options and the implications of their child *not* being included in the assessment. | Mean = 4.69<br>SD = 0.70<br>N = 217 | Mean = 3.63**<br>SD = 1.08<br>N = 217 | Mean = 2.22<br>SD = 1.22<br>N = 217 |
| 14. | Decisions about participation are documented on the student's IEP or on an additional form that is attached to the IEP. | Mean = 4.52<br>SD = 1.08<br>N = 218 | Mean = 4.37<br>SD = 1.04<br>N = 217 | Mean = 3.39**<br>SD = 1.61<br>N = 219 |
| 15. | Decision about accommodations is made by a person (or group of persons) who knows the student. | Mean = 4.92<br>SD = 0.36<br>N = 219 | Mean = 4.63<br>SD = 0.65<br>N = 219 | Mean = 3.80**<br>SD = 1.24<br>N = 219 |
| 16. | Decision about accommodations is based on the student's current level of functioning and learning characteristics. | Mean = 4.87<br>SD = 0.52<br>N = 217 | Mean = 4.50<br>SD = 0.74<br>N = 216 | Mean = 3.60**<br>SD = 1.28<br>N = 217 |
| 17. | A form is used that lists the variables to consider in making the accommodation decision and that documents the decision for each student and reasons for it. | Mean = 4.55<br>SD = 0.81<br>N = 218 | Mean = 4.12<br>SD = 0.97<br>N = 218 | Mean = 2.43<br>SD = 1.46<br>N = 216 |
| 18. | Accommodation guidelines link instructional accommodations to decisions about assessment accommodations. | Mean = 4.65<br>SD = 0.73<br>N = 216 | Mean = 4.11<br>SD = 1.01<br>N = 217 | Mean = 2.70<br>SD = 1.39<br>N = 217 |
| 19. | Decision about accommodations is *not* based on the program setting. | Mean = 4.50<br>SD = 1.06<br>N = 219 | Mean = 4.18<br>SD = 1.08<br>N = 218 | Mean = 3.32**<br>SD = 1.30<br>N = 217 |
| 20. | Decision about accommodations is *not* based on the category of disability. | Mean = 4.48<br>SD = 1.08<br>N = 218 | Mean = 4.23<br>SD = 1.05<br>N = 215 | Mean = 3.39**<br>SD = 1.29<br>N= 214 |
| 21. | Decision about accommodations is *not* based on the percent time in the mainstream classroom. | Mean = 4.43<br>SD = 1.09<br>N = 217 | Mean = 4.22<br>SD = 1.05<br>N = 215 | Mean = 3.35**<br>SD = 1.35<br>N = 214 |
| 22. | Decisions about accommodations are documented on the student's IEP or on an additional form that is attached to the IEP. | Mean = 4.65<br>SD = 0.91<br>N = 215 | Mean = 4.44<br>SD = 0.94<br>N = 215 | Mean = 3.49**<br>SD = 1.46<br>N = 216 |

**Table 2. Means, Standard Deviations, and Number of Respondents for each Item (continued)**

| | Criteria | Desirability | Feasibility | Actual Use |
|---|---|---|---|---|
| 23. | Parents understand the accommodation options and the implications of their child using an accommodation in assessment. | Mean = 4.78<br>SD = 0.61<br>N = 215 | Mean = 3.94<br>SD = 1.05<br>N = 215 | Mean = 2.71<br>SD = 1.22<br>N = 216 |
| 24. | A written policy exists about who is included when calculating participation or exclusion rates. | Mean = 4.48<br>SD = 1.03<br>N = 215 | Mean = 4.18<br>SD = 1.09<br>N = 213 | Mean = 2.70<br>SD = 1.50<br>N = 210 |
| 25. | Rates of exclusion that are specific to students with disabilities, and reasons for the exclusion, are reported when assessment results are reported. | Mean = 4.38<br>SD = 1.07<br>N = 212 | Mean = 4.03<br>SD = 1.18<br>N = 207 | Mean = 2.40<br>SD = 1.49<br>N = 199 |
| 26. | Data reports include information from all test takers. | Mean = 4.51<br>SD = 0.94<br>N = 204 | Mean = 4.31<br>SD = 1.03<br>N = 203 | Mean = 3.09<br>SD = 1.54<br>N = 199 |
| 27. | Records are kept so that data for students with disabilities could be reported separately, overall, or by other breakdowns. | Mean = 4.48<br>SD = 1.06<br>N = 212 | Mean = 4.22<br>SD = 1.07<br>N = 210 | Mean = 2.80<br>SD = 1.51<br>N = 208 |
| 28. | Records are kept of the use of accommodations by students with disabilities, by type of accommodation, so that the information could be reported either by individual student or in aggregate. | Mean = 4.27<br>SD = 1.18<br>N = 212 | Mean = 3.80**<br>SD = 1.29<br>N = 210 | Mean = 2.33<br>SD = 1.47<br>N = 208 |
| 29. | Parents understand the reporting policy for their child's assessment results. | Mean = 4.70<br>SD = 0.74<br>N= 213 | Mean = 3.84**<br>SD = 1.13<br>N = 212 | Mean = 2.28<br>SD = 1.09<br>N = 211 |

* Means for desirability and feasibility are significantly different from each other in every criteria except Criteria 6.

** Mean is significantly different from the average (Desirability M = 4.49; Feasibility M = 4.08; Actual Use M = 2.91). (See Seyfarth, 1998, for specific $t$ values for all statements.)

- Criterion 11, "Decision guidelines recognize that only a small percentage of students with disabilities need to participate in an alternate assessment (e.g., those with severe disabilities, <1% of all students)," ($t(212) = -4.83$, $p = 0.000$).

### Feasibility

To assess the perceived feasibility of the criteria, one-sample t-tests were conducted and compared to the overall sample mean for the feasibility of the criteria ($M = 4.08$). Again, the significance level was set at 0.005. This analysis resulted in six criteria that were significantly below the mean for feasibility. Table 2 contains the full results of the comparisons. The criteria are:

- Criterion 1, "All students, including students with disabilities, are to participate in state and district assessments," ($t(220) = -9.34$, $p = 0.000$).
- Criterion 6, "A student must participate in an assessment if the student receives *any* instruction on the content assessed, regardless of where instruction occurs," ($t(212) = -11.37$, $p = 0.000$).
- Criterion 12, "Parents understand the state/district accountability system," ($t(218) = -8.45$, $p = 0.000$).
- Criterion 13, "Parents understand participation options and the implications of their child *not* being included in the assessment," ($t(217) = -6.03$, $p = 0.000$).
- Criterion 28, "Records are kept of the use of accommodations by students with disabilities, by type of accommodation, so that the information could be reported either by individual student or in aggregate," ($t(210) = -3.10$, $p = 0.002$).
- Criterion 29, "Parents understand the reporting policy for their child's assessment results," ($t(212) = -3.03$, $p = 0.003$).

### Comparison of Desirability and Feasibility

Paired-samples t-tests were conducted in order to examine the question of whether some criteria were seen as less feasible than they were desirable. The significance level was set at 0.005. The ratings for all statements but one were found to be significantly different in desirability and feasibility. The only criterion that was not rated to be significantly different in desirability and feasibility was Criterion 6, "A student must participate in an assessment if the student receives *any* instruction on the content assessed, regardless of where instruction occurs," ($t(210) = 2.39$, $p = 0.018$). The means for desirability and feasibility of this criterion were 3.21 and 3.03, respectively.

## Actual Use

To assess which criteria were most frequently used by special education directors, one-sample t-tests were conducted on the reported actual use of the criteria and compared to the overall sample mean for actual use (M = 2.91). Significance level was again set at 0.005. This analysis resulted in twelve criteria that were significantly above the mean for actual use. They are:

- Criterion 2, "The decision about participation is made by a person (or a group of people) who knows the student," (t (221) = 12.34, p = 0.000).
- Criterion 3, "The decision about participation is based on the student's current level of functioning and learning characteristics," (t (217) = 8.56, p = 0.000).
- Criterion 7, "Decision about participation is not based on the program setting," (t (215) = 3.26, p = 0.001).
- Criterion 8, "Decision about participation is *not* based on the category of disability," (t (217) = 3.89, p = 0.000).
- Criterion 9, "Decision about participation is *not* based on the percentage of time in the mainstream classroom," (t (215) = 3.23, p = 0.001).
- Criterion 14, "Decisions about participation are documented on the student's IEP or on an additional form that is attached to the IEP," (t (219) = 4.48, p = 0.000).
- Criterion 15, "Decision about accommodations is made by a person (or group of persons) who knows the student," (t (219) = 10.80, p = 0.000)
- Criterion 16, "Decision about accommodations is based on the student's current level of functioning and learning characteristics," (t (217) = 8.07, p = 0.000).
- Criterion 19, "Decision about accommodations is *not* based on the program setting," (t (217) = 4.75, p = 0.000).
- Criterion 20, "Decision about accommodations is *not* based on the category of disability," (t (214) = 5.56, p = 0.000).
- Criterion 21, "Decision about accommodations is *not* based on the percent time in the mainstream classroom," (t (214) = 4.82, p = 0.000).
- Criterion 22, "Decisions about accommodations are documented on the student's IEP or on an additional form that is attached to the IEP," (t (216) = 5.87, p = 0.000).

## Discussion

Accountability systems, and statewide assessments as a part of those systems, have become commonplace in the United States. As they have become commonplace, we have started to identify issues with which many states are struggling. One such issue is decisions about participation of students with disabilities in state or district assessments. Kentucky is currently the only state that mandates that all students must participate in their state accountability system (Thurlow et al., 1997). Other states have a variety of rules about how they decide who participates, what type of accommodations are available to students with disabilities, and how they will report the results of those students' tests (Thurlow et al., 1997).

Recent changes in laws (IDEA, Title I) have resulted in federal mandates that states must begin to account for the educational progress of all students, including those with disabilities (Individuals with Disabilities Education Act 1997 Amendments). Based on these changes, states will have to move toward including more students in their assessments, and to devise alternate assessments for students for whom the general assessments are inappropriate. While policy mandates are important in moving a field forward, it is also important to examine stakeholder support for the mandates. Such support is critical to having them implemented as intended.

The results of the survey of stakeholders' perceptions of best practice criteria for including students with disabilities in large-scale assessments demonstrate that while there is overall agreement with the suggested criteria, stakeholders were hesitant about certain issues. Nearly all thought criteria were less feasible than they were desirable, and they reported using certain practices significantly more than others. More specifically, there was less agreement on the criteria that restricted individuals' abilities to decide about participation of students with disabilities. For example, the three criteria that educators reported to be significantly less desirable than the rest were those that made broad, absolute statements about the participation of students with disabilities in assessment. An example of such a criterion is the first criterion on the survey, "All students, including students with disabilities, are to participate in state and district assessments."

There are a number of possible reasons for why educators might find the broad, global statements about participation less desirable than the other criteria. One possible reason is that educators might want to make those decisions on an individual basis. They may believe that there are individual differences among students that affect the appropriateness of the test for them, and that we should not restrict educators' abilities to make individual decisions for students. It is also possible that some educators may simply think that it is a bad idea to include all students in these tests, that it is not worth the time taken to test students with disabilities, that the students will bring down scores, that the students cannot achieve high standards, or that students with disabilities should be kept separate from the mainstream of education.

When looking at the perceived feasibility of the criteria, the least desirable criteria were also generally indicated to be less feasible than the other criteria. However, there were some criteria that were rated low in feasibility but not rated low in desirability. Three of the four criteria that fit this description had to do with parents. For example, Criterion 13 states, "Parents understand participation options and the implications of their child *not* being included in the assessment." The fourth criterion that fit this description had to do with record-keeping, "Records are kept of the use of accommodations by students with disabilities, by type of accommodation, so that the information could be reported by individual student or in aggregate."

While educators believe that it is important for parents to understand the accountability system and the implications for their children of taking part in the system, they see it as less possible than the rest of the criteria. This is a disturbing belief, given all that we know about the importance of parent participation in their children's education (e.g., Grolnick, Benjet, Kurowski, & Apostoleris, 1997; Keith, Keith, Troutman, Bickley, Trivette, & Singh, 1993). There are a number of possible reasons for this. It may be that the systems, as designed currently, are so complex that educators feel unprepared to educate parents about them. It could also be that this belief is part of educators' general mistrust of parent involvement (e.g., Grolnick et al., 1997). Educators' difficulty with the statements could also be partially explained by the way they are worded, that they always start with the phrase "Parents understand." This may imply to educators that this criterion is only met when all parents understand. Educators might have reacted more favorably if the criterion had read "most parents understand," or something less global. Another possible reason for this difference could be skepticism on educators' part about parents' ability to understand the systems.

Keeping records of provided accommodations was also seen as significantly less feasible than the other criteria. It may be that states are so new at implementing these systems that they have no routine mechanism in place for recording the accommodations provided. Another possible reason for the low feasibility may be that including students with disabilities in accountability systems is so new to educators that they are overwhelmed by the amount of work involved and to require them to do anything they are not currently doing seems to be too much. Another reason that this was seen as low in feasibility may be the complexity involved in accommodations. It may be difficult for teachers to decide when an accommodation is made (e.g., if a student sits in a specific part of the room, is that an accommodation?), or to be sure that the accommodation was actually implemented (e.g., if the IEP team decides that a student can use a calculator on the assessment, but it is not necessarily clear that the student actually used the calculator—was the accommodation implemented?).

Criterion 6 suggests that students must participate in the assessment if they receive any relevant instruction, regardless of where it occurs. This was the only criterion on which the rated desirability and feasibility of the criterion was not significantly different. This was likely because

the perceived desirability and feasibility of the criterion was extremely low, as compared to the other criteria. This criterion may have been rated so low by educators because of a concern over curricular validity of the instrument. In other words, is it appropriate and valid to test a student on information the student has not received? A possible scenario would be if an eighth-grade student with severe cognitive disabilities has content called "mathematics," in which he is taught to recognize the numbers one and two. Would it be appropriate to measure his achievement in mathematics on an eighth-grade test? A way this concern might be addressed would be to modify the wording of the criterion to indicate that students are expected to be assessed, by alternate or regular assessments, if they receive any instruction on the content in the assessment. This could give the flexibility to address the scenario above, while also ensuring that students with disabilities are a part of the ongoing accountability system.

The ratings of actual use contrast with the ratings of desirability and feasibility. No criterion that was significantly higher than the mean for actual use (M = 2.91) was rated to be low in desirability or feasibility. The criteria that were rated to be higher than the mean for actual use were generally focused either on making individual decisions for students or on documentation of the decision (e.g., decision made by those who know the student, decision is *not* based on program setting). Again, this seems to support educators' preference for making individualized decisions for students, and it also implies that states are currently documenting the decisions they are making for individual students.

Findings from this study provide a valuable first step in understanding educators' thoughts on including students with disabilities in large-scale assessments. From this study, we can have a better understanding of where there is likely to be agreement with the policies states are implementing, and where there may be problems getting stakeholder support. It will be important for policymakers to understand that while there is widespread agreement with most of the criteria, there is also universal pessimism about the feasibility of actually implementing the criteria.

One of the principal limitations of this study was sampling methods. States were not chosen randomly within the census regions, and participants were not chosen randomly in most states. This could have resulted in a biased sample that would be more predisposed to agree with NCEO's criteria.

Another possible limitation of the study was the method for distributing surveys to teachers and other IEP team members. Asking the district special education administrator to disseminate them could again lead to biased sampling methods, as perhaps the administrator provided surveys to staff who had strong opinions about this material. It is also possible that the staff could have felt pressure to agree with the criteria, given the fact that it was handed to them by their supervisor.

This study also focused on educators, both administrators and direct service providers, as the stakeholders in making these decisions. Even more information could have been obtained had

parents and students been included in the stakeholder population. The choice to include only educators allowed us to focus on those actually implementing the policies, but moved our focus away from the clients of education—students and their families.

## Policy Implications and Future Directions

It appears that while educators do see the NCEO criteria as desirable, there is significantly less confidence in the implementation of the criteria. This is a concern, given the new guidelines in the IDEA 1997 Amendments that require states to include students with disabilities in their tests, and also to report on their participation.

The results of the survey do fit with results from other analyses of national trends. Thurlow et al. (1997) found that while most states do have participation and accommodation policies in place or in development, the policies often provide less than comprehensive suggestions about making these decisions. IEP teams are usually, but still not always, the decision-making body about who will participate in the tests. Additionally, only one state, Kentucky, requires that all students participate in its statewide accountability system, and in Kentucky the IEP team makes a decision regarding whether the student participates in the regular assessment, the regular assessment with accommodations, or in an alternate assessment (Thurlow et al., 1997).

Educators will continue to struggle with these issues over time, given the recent legislative changes, and also the trends nationally for participation of students with disabilities (Thurlow et al., 1997). It seems that what is needed most at this point is assistance to the states, and through them to local educators, in how to implement some of these changes. Given the information gathered in this survey, it also will be necessary to find a way to implement the changes without restricting too severely educators' decision-making abilities.

Another need made clear by this survey is assistance to educators and parents in helping parents understand these systems and their children's place in the systems. Need for technical assistance to states, including teacher inservice training, as well as preservice training on this issue is apparent.

Research on states that are most successfully implementing these changes would be helpful in determining how to encourage the participation of students with disabilities in state or district assessment without restricting educator's decision-making freedoms. Additionally, research on the effects of accommodations is desperately needed. Finally, research on effective technical assistance to the states would be helpful—what do states need to implement the changes that are occurring?

As we move toward full participation for all students in accountability systems, there will, no doubt, be difficulties. We need to keep in mind the basic question, for whom do we hold our

schools accountable? And, within that, are students with disabilities provided the same rights as other students within our educational system? If so, they have the right to be included in the systems intended to ensure educational quality.

## References


Algozzine, B. (1993). Including students with disabilities in systemic efforts to measure outcomes: Why ask why? In J. E. Ysseldyke & M. L. Thurlow (Eds.), *Views on inclusion and testing accommodations for students with disabilities* (pp. 5-10). (Synthesis Report 7). Minneapolis, MN: National Center on Educational Outcomes.

Allington, R. L., & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York. *Educational Policy, 6,* 397-414.

Almond, P., Tindal, G., & Stieber, S. (1997). *Linking inclusion to conclusions: An empirical study of participation of students with disabilities in statewide testing programs* (Oregon Report 1). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington DC: American Psychological Association.

Baldwin, R. S., Murfin, P., Ross, G. R., Seidel, J., Clements, N., & Morris, C. (1989). Effects of extending administration time on standardized reading achievement tests. *Reading Research and Instruction, 29,* 33-38.

Bennett, R. E., Rock, D. A., & Kaplan, B. A. (1987). SAT differential item performance for nine handicapped groups. *Journal of Educational Measurement, 24,* 41-55.

Bond, L., Braskamp, D., & Roeber, E. (1996). *The status of student assessment programs in the United States*. Oak Brooks, IL: NCREL.

Elliott, J., Thurlow, M., & Ysseldyke, J. (1996). *Assessment guidelines that maximize the participation of students with disabilities in large-scale assessments: Characteristics and considerations* (Synthesis Report 25). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Erickson, R., & Thurlow, M. (1997). *1997 state special education outcomes*. Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Erickson, R., Thurlow, M., Thor, K., & Seyfarth, A. (1996). *State special education outcomes, 1995*. Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Erickson, R., Thurlow, M., & Ysseldyke, J. (1996). *Neglected numerators, drifting denominators, and fractured fractions: Determining participation rates for students with disabilities in statewide assessment programs* (Synthesis Report 23). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Gilman, C. J. (1994). Stakeholder reactions to developing a model for early childhood. *Special Services in the Schools, 9,* 173-180.

Grolnick, W. S., Benjet, C., Kurowski, C. O., & Apostoleris, N. H. (1997). Predictors of parent involvement in children's schooling. *Journal of Educational Psychology, 89,* 538-548.

Henry, G. T., Dickey, K. C., & Areson, J. C. (1991). Stakeholder participation in educational performance monitoring systems. *Educational Evaluation and Policy Analysis, 13,* 177-188.

Jayanthi, M., Epstein, M. H., Polloway, E. A., & Bursuck, W. D. (1996). A national survey of general education teachers' perceptions of testing adaptations. *The Journal of Special Education, 30,* 99-115.

Keith, T. Z., Keith, P. B., Troutman, G. C., Bickley, P. G., Trivette, P. S., & Singh, K. (1993). Does parental involvement affect eighth-grade student achievement? Structural analysis of national data. *School Psychology Review, 22,* 474-496.

McDonnell, L. M., McLaughlin, M. J., & Morison, P. (Eds.) (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington D. C.: National Academy Press.

McGill-Franzen, A., & Allington, R. L. (1993). Flunk 'em or get them classified: The contamination of primary grade accountability data. *Educational Researcher, 22,* 19-22.

Phillips, S. E. (1993). Testing condition accommodations for disabled students. *West's Education Law Quarterly, 2,* 366-89.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education, 7,* 93-120.

Ragosta, M., & Kaplan, B. A. (1986). *A survey of handicapped students taking special test administrations of the SAT and GRE.* (Report number 5, Studies of admissions testing and handicapped people). Princeton, NJ: Educational Testing Service.

Renihan, P. J., & Renihan, F. I. (1994). Institutionalizing school improvement: A recipe for success. *NASSP Bulletin, 78 (559),* 84-92.

Rock, D. A., Bennett, R. E., & Kaplan, B. A. (1987). Internal construct validity of a college admissions test across handicapped an nonhandicapped groups. *Educational and Psychological Measurement, 47,* 193-205.

Seyfarth, A. L. (1998). *Perceived desirability and feasibility of specific criteria for large-scale assessment and accountability systems.* Unpublished master's thesis, University of Minnesota, Minneapolis.

Shriner, J. G., & Thurlow, M. L. (1992). *State special education outcomes, 1991*. Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Siskind, T. G. (1993). Modifications in statewide criterion-referenced testing programs to accommodate pupils with disabilities. *Diagnostique, 18*, 233-249.

Thurlow, M., Hurley, C., Spicuzza, R., & El Sawaf, H. (1996). *A review of the literature on testing accommodations for students with disabilities* (Minnesota Report 9). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Thurlow, M. L., Langenfeld, K., Nelson, J. R., Shin, H., & Coleman, J. (1998). *State accountability reports: What do they say about students with disabilities?* (Technical Report 20). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Thurlow, M. L., Seyfarth, A. L., Scott, D. L., & Ysseldyke, J. E. (1997). *State assessment policies on participation and accommodations for students with disabilities: 1997 update* (Synthesis Report 29). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Thurlow, M. L., Ysseldyke, J. E., & Silverstein, B. (1993). *Testing accommodations for students with disabilities: A review of the literature*. (Synthesis Report 4). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Ysseldyke, J. E., Thurlow, M. L., McGrew, K. S., & Shriner, J. G. (1994). *Recommendations for making decisions about the participation of students with disabilities in statewide assessment programs.* (Synthesis Report 15). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Zlatos, B. (1994). Don't test, don't tell: Is 'academic red-shirting' skewing the way we rank our schools? *The American School Board Journal, 181 (11)*, 24-28.

## Appendix A

Markers for the Participation, Accommodation, and Reporting Criteria

**Markers for Criterion 1: Participation**

1. All students, including students with disabilities, are to participate in state and district assessments
2. The decision about participation is made by a person (or a group of people) who knows the student
3. The decision about participation is based on the student's current level of functioning and learning characteristics.
4. A form is used that lists the variables to consider in making participation decisions.
5. Reason(s) for exclusion is documented.
6. A student must participate in an assessment if the student receives any instruction on the content assessed, regardless of where instruction occurs.
7. Decision about participation is not based on the program setting.
8. Decision about participation is not based on the category of disability.
9. Decision about participation is not based on the percentage of time in the mainstream classroom.
10. Decision guidelines allow for some students to participate in an alternate assessment or, when appropriate, in part of an assessment or assessment procedure.
11. Decision guidelines recognize that only a small percentage of students with disabilities need to participate in an alternate assessment (e.g., those with severe disabilities, <1% of all students).
12. Parents understand the state/district accountability system.
13. Parents understand participation options and the implications of their child not being included in the assessment.
14. Decisions about participation are documented on the student's IEP or on an additional form that is attached to the IEP.

**Markers for Criterion 2: Accommodations**

1. Decision about accommodations is made by a person (or group of persons) who knows the student.
2. Decision about accommodations is based on the student's current level of functioning and learning characteristics.

3. A form is used that lists the variables to consider in making the accommodation decision and that documents the decision for each student and reasons for it.
4. Accommodation guidelines link instructional accommodations to decisions about assessment accommodations.
5. Decision about accommodations is not based on the program setting.
6. Decision about accommodations is not based on the category of disability.
7. Decision about accommodations is not based on the percent time in the mainstream classroom.
8. Decisions about accommodations are documented on the student's IEP or on an additional form that is attached to the IEP.
9. Parents understand the accommodation options and the implications of their child using an accommodation in assessment.

**Markers for Criterion 3: Reporting**

1. A written policy exists about who is included when calculating participation or exclusion rates.
2. Rates of exclusion that are specific to students with disabilities, and reasons for the exclusion, are reported when assessment results are reported.
3. Data reports include information from all test takers.
4. Records are kept so that data for students with disabilities could be reported separately, overall, or by other breakdowns.
5. Records are kept of the use of accommodations by students with disabilities, by type of accommodation, so that the information could be reported either by individual student or in aggregate.
6. Parents understand the reporting policy for their child's assessment results.

## Appendix B

Copy of Survey Instrument

Name: _____

Position/Title: _____

State: _____

Approximate Number of Students in District: _____

Type of District: Urban Suburban Rural Other:_____

Directions: Please rate each assessment practice listed below on a 1-5 scale (1 = low; 5 = high) in terms of its DESIRABILITY (good idea?), its FEASIBILITY (how practical it would be to actually implement the practice in your state or district assessments), and ACTUAL USE (what your state or district has presently implemented). There are no right answers. We are interested in what you think.

| | Desirability | | | | | Feasibility | | | | | Actual Use | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LOW | | | | HIGH | LOW | | | | HIGH | LOW | | | | HIGH |
| Participation of students with disabilities in state and district assessments | | | | | | | | | | | | | | | |
| 1. All students, including students with disabilities, are to participate in state and district assessments. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 2. The decision about participation is made by a person (or a group of people) who knows the student. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 3. The decision about participation is based on the student's current level of functioning and learning characteristics. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 4. A form is used that lists the variables to consider in making participation decisions. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 5. Reason(s) for exclusion is documented. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 6. A student must participate in an assessment if the student receives <u>any</u> instruction on the content assessed, regardless of where instruction occurs. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 7. Decision about participation is <u>not</u> based on the program setting. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 8a Decision about participation is <u>not</u> based on the category of disability. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 9a Decision about participation is <u>not</u> based on the percentage of time in the mainstream classroom. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |

(Continued on next page)

| | | Desirability | | | | | Feasibility | | | | | Actual Use | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LOW | | | | HIGH | LOW | | | | HIGH | LOW | | | | HIGH |
| 8b. | Decision guidelines allow for some students to participate in an alternate assessment or, when appropriate, in part of an assessment or assessment procedure. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 9b. | Decision guidelines recognize that only a small percentage of students with disabilities need to participate in an alternate assessment (e.g., those with severe disabilities, <1% of all students). | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 10. | Parents understand the state/district accountability system. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 11. | Parents understand participation options and the implications of their child not being included in the assessment. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 12. | Decisions about participation are documented on the student's IEP or on an additional form that is attached to the IEP. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Accommodations in Testing | | | | | | | | | | | | | | | | |
| 13. | Decision about accommodations is made by a person (or group of persons) who knows the student. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 14. | Decision about accommodations is based on the student's current level of functioning and learning characteristics. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 15. | A form is used that lists the variables to consider in making the accommodation decision and that documents the decision for each student and reasons for it. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 16. | Accommodation guidelines link instructional accommodations to decisions about assessment accommodations. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 17. | Decision about accommodations is not based on the program setting. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |

(Continued on next page)

BEST COPY AVAILABLE

| | Desirability | | | | | Feasibility | | | | | Actual Use | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LOW | | | | HIGH | LOW | | | | HIGH | LOW | | | | HIGH |
| 18. Decision about accommodations is <u>not</u> based on the category of disability. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 19. Decision about accommodations is <u>not</u> based on the percent time in the mainstream classroom. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 20. Decisions about accommodations are documented on the student's IEP or on an additional form that is attached to the IEP. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 21. Parents understand the accommodation options and the implications of their child using an accommodation in assessment. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| <u>Reporting of Results</u> | | | | | | | | | | | | | | | |
| 22. A written policy exists about who is included when calculating participation or exclusion rates. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 23. Rates of exclusion that are specific to students with disabilities, and reasons for the exclusion, are reported when assessment results are reported. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 24. Data reports include information from all test takers. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 25. Records are kept so that data for students with disabilities could be reported separately, overall, or by other breakdowns. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 26. Records are kept of the use of accommodations by students with disabilities, by type of accommodation, so that the information could be reported either by individual student or in aggregate. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 27. Parents understand the reporting policy for their child's assessment results. | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |

(Continued on next page)

BEST COPY AVAILABLE

Additional Comments:

Thank you!

The College of Education
& Human Development

UNIVERSITY OF MINNESOTA